Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 12 (2012) 398 – 403

Procedia Computer Science

Complex Adaptive Systems, Publication 2
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2012- Washington D.C.

Security Through Behavioral Biometrics and Artificial Intelligence

Benjamin Purgason[a], David Hibler[b]*

[a,b]*Christopher Newport Unviersity, PCSE Department, 1 University Place, Newport News, Virginia 23606, USA*

Abstract

The purpose of this paper is to validate a method of collecting and analyzing behavioral biometric data in order to authenticate a user's identity. The method uses the time to transition from one finger to another while typing, a form of Key Interval Time biometrics (KIT). The analysis is performed using feed-forward neural nets. The User Rights and Integrity Enforcement Logic platform, URIEL, was developed to implement this methodology. The URIEL system analyzes KIT data in near-real-time. Its purpose is to demonstrate an effective means of authenticating and distinguishing amongst a group of privileged users. This concept is being assessed through live human trials. The first trial to be completed was an anonymous trial whose volunteer participants were student members of the Physics Computer Science and Engineering Department of Christopher Newport University. The high level of uniformity in this population increased the difficulty in distinguishing between individuals thereby simulating the URIEL platform's intended purpose: the regulation and identification of a small group of privileged users. The results of the first trial were encouraging and are presented here along with a discussion of the impact of variations of the methodology on the results.

Keywords: Biometrics; Neural Nets; Security

1. Introduction

There is a growing need to find a new means to authenticate a user's identity. We propose the utilization of artificial intelligence and behavioral biometrics in order to identify users based upon their actions, tendencies, and idiosyncrasies. Specifically, we propose the analysis of an attribute that is difficult to willfully and precisely alter simultaneously: the time required to transition from one finger to another while typing, a form of Key Interval Time biometrics (KIT). The User Rights and Integrity Enforcement Logic platform, URIEL, was developed to implement this methodology using neural nets.

The purpose of this paper is to validate the concept of analyzing behavioral biometric KIT data through the use of artificial intelligence techniques as an effective means of identifying and distinguishing humans from one another.

* Corresponding author. Tel.: 757-594-7360; fax: 757-504-7919.

 doi:10.1016/j.procs.2012.09.093

An additional purpose is to assess the viability of analyzing behavioral biometric data in near-real-time via feed-forward neural nets as an effective means of authenticating and distinguishing amongst a group of privileged users.

1.1. Background

The field of behavioral biometrics has seen many papers published discussing possible means of authentication, typically augmenting the traditional password based authentication mechanism. Though many options have been explored, few have been found to be sufficiently effective as to be deployed on a wide scale [1].

A large study, Bergadano's "Identity verification through dynamic keystroke analysis" was conducted using 130 participants while recording information regarding the rhythm of their typing [2]. Using this collected data Bergadano et al. were capable of identifying imposters as well as differentiating between individuals. Though digraph comparison of KIT data rather than a neural network based solution was used, this work lends credibility to the concept of scrutinizing KIT based keyboard dynamics to determine a user's identity.

The study by Gianvecchio, et al. [3] used neural networks. The networks were able to differentiate between the input patterns produced by a human and those produced by a bot thereby showing the ability to create a general profile of humans. Using neural networks, meaningful patterns were successfully extracted from the inputs provided by a human showing the potential existence of distinguishing patterns being embedded in such input. On the whole this study, while sharing many commonalities with this paper, is fundamentally different in that its objective was to detect bots posing as human game players whereas the objective of this paper is to differentiate between multiple humans who may or may not be attempting to impersonate one another through the use of KIT based keyboard dynamics.

Ngugi's "Typing Biometrics: Impact of Human Learning on Performance Quality", [4] is perhaps the most significant paper in terms of support for the concept behind the URIEL platform. Ngugi, like Lee et al. [5], raises a key point that can lead to reduced identification accuracy when keyboard dynamics are utilized: typing patterns in individuals change over time. Specifically, any amount of time allowing a human to better learn a repeatable pattern will alter his or her typing patterns. This is not expected to negatively affect the accuracy of the neural net used in this research since the net will be retrained after each successful identification and take newly available data into account. The retraining will slowly adapt a person of interest's profile over time and adjust for shifts in typing patterns. This suggests that the usage of a neural net will, through its innate characteristics, mitigate or completely solve the difficulties noted by Ngugi et al.'s study.

The second section of this paper describes the URIEL methodology in detail. The third section considers tests of this implementation and compares variations in the methodology. Finally, the last section provides conclusions and suggestions for further work

2. URIEL Methodology

2.1. Human Measurement Metrics

Human Measurement Metrics (HMMs) are the attributes measured in order to establish the identity of humans and draw distinctions between them. The HMMs used for this are the coordination levels of individual fingers. Due to the nature of the HMMs and the large variation in available hardware from site to site the HMMs are measured utilizing minimal hardware.

Firstly, to overcome the wide variation in available hardware the keyboard was selected as the sole piece of hardware to be used in the collection of raw input data.

The keyboard was divided into ten "key-groups" that roughly correspond to the responsibilities of individual fingers when using proper technique on a keyboard implementing the QWERTY layout. The groups are numbered from 0 through 9. . Groups zero through seven contain letters and are as follows: group 0 - Q, A, Z; group 1 - W, S, X; group 2 - E, D, C; group 3 - R, T, F, G, V, B; group 4 - Y, U, H, J, N, M; group 5 - I, K; group 6 - Q, L; and group 7 - P. All groups that have a letter as a member contain both the upper-case and lower-case variant on that letter. Key group eight contains the spacebar and the final group, group 9, contains the remaining keys. Group 9

was created in order to better assist a neural network in removing potentially noisy information that could be disruptive to its ability to learn. Group 9 is considered to be noisy because it contains all of the characters on a keyboard that require extreme reach, movement of the hand as a whole, or multiple keys being depressed simultaneously to create.

During typing all of the transitions that occur and the time it took to complete them are recorded, including those where transitions that originate and end within the same group. At the conclusion of the user observation all of the entries for a given transition are averaged before being saved for later use. In this manner a single data set is created containing exactly one entry per possible transition after each observation of a user and is saved into a ".raw" data file. Each of these raw files contains one-hundred entries, one for each possible transition, effectively a ten by ten matrix of average transition times. Once enough raw files are created an URIEL module consolidates them to create profiles. These files are simply for improving the portability of the data and no processing occurs on the actual data contained within.

2.2. Neural Network

Each complete dataset of 100 average transition times contained in the profile files was analyzed via multiple feed-forward, back-propagation neural networks. A momentum parameter of 0.6 was utilized to allow the neural networks to more easily avoid local minima and accelerate the rate of training. The profiles used for training contained 100 possible transition times, each taken as an input for 100 total inputs, as well as the 2 expected output values. The configuration for each network was 100 input layer neurons, 12 hidden layer neurons, and 2 output layer neurons. The training process was only performed using the designated training sets. The remaining data sets were not shown to the neural networks until the testing phase to ensure learning was successfully completed.

For the implementation used in this study, one neural network was created for each user and trained to recognize exactly one unique user. All available training sets were used by all neural nets, positive reinforcement if the data set was created by the person of interest to a given neural net otherwise negative reinforcement was used. Training times were less than two minutes under most circumstances.

3. Human Trials

The first human trial was an anonymous trial whose volunteer participants were student members of the Physics Computer Science and Engineering Department of Christopher Newport University. No compensation was offered other than the chance to assist in validating a methodology for the identification and classification of humans.

The population of study participants contained a limited amount of variety. The nominal participant was between the ages of nineteen and twenty-four, an undergraduate, and studying a technology related field. The decision to encourage such uniformity within the first human trial population was two-fold. Firstly, it stands to reason that similar people should be harder to distinguish from one another than more dissimilar people. The high level of uniformity increases the difficulty in distinguishing between individuals thereby simulating the URIEL platform's intended purpose: the regulation and identification of a small group of privileged users. Secondly, this high-level of population uniformity allows for its effects to be measured through subsequent human trials.

Each participant in the trial typed a standard passage from a poem. Transitions were measured with one millisecond accuracy and were then averaged. That is to say all recorded instances of transition 0 to 0 were averaged and saved. The process was repeated for all remaining transitions. This resulted in 100 data points of averaged identity data for each entry in the user profile. Users typed the same passage on different occasions to provide multiple entries for their profile.

Even though averaging should have a mitigating or standardizing effect there were significant differences when the averaged identity data from one user was compared to the data from other users. When each of the 100 data points, corresponding to the averages for each of the 100 possible transitions, from the averaged identity data were compared across all of the study participants it was noted that many of these had variances greater than 4000 and there were very large differences in variation between one data point and the next. Even so a statistically significant correlation was found between all except one pair of user's averaged identity data meaning a higher than 0.90 correlation coefficient. The minimum correlation between any two users was 85% while all other combinations were greater than 87%; the average correlation between a pair of user profiles was 91%.

This high level of correlation between the various users' averaged identity data indicates the neural networks may not be learning the relationship between two transition times but rather they may be learning the transition times themselves or perhaps taking both into account while giving priority to the transition times themselves.

During testing URIEL was permitted to either make a decision or to refuse to decide based upon its level of confidence in its answer. Instances wherein URIEL refused to answer combined with instances of correctly rendered decisions are regarded as "Fail Secure Acceptable" (FSA) since in this instance URIEL did not decide wrongly and would simply require additional data to come to a decision. Instances where URIEL did make a decision and was correct in its assertion are regarded as URIEL making an "authoritative decision" since it made a decision with a high degree of confidence. Finally, instances where URIEL rendered a decision but chose wrongly were regarded simply as "failures" since URIEL performed unacceptably regardless of the situation.

Initial results of the first human trial were very promising. URIEL was able to make an authoritative decision in 82.178% of all test cases provided, refused to make a decision in 17.822% of all cases, and did not choose wrongly. This particular set of data used in testing URIEL included more than one profile with a significantly lower number of entries, profiles from participants who did not provide the full quantity of datasets requested. After rerunning tests on the data collected from the human trial, typical expectations of URIEL performance may be set as an 82% rate of correct decisions rendered, a 16% rate of refused decisions, and finally a 2% rate of insecure failure.

The number of entries contained within each of the profiles had a substantial impact on the overall performance of URIEL. Each additional entry within the profiles conferred a substantial improvement over the last up to a point. This was not a linear benefit but instead one of decreasing returns. Up until each profile contained about twelve entries each entry within the profile improved the overall performance of URIEL's accuracy. After the twelfth profile entry the performance of URIEL rose to a greater than eighty percent accuracy rate. It is at this point that additional profiles did not appear to have a significant effect. This limitation is not expected to be a hard limit but one induced by the implementation used. Through the optimization of the underlying neural networks and URIEL it could be possible to improve this limitation.

3.1. URIEL Decision Making Strategies

The decision making process utilized by URIEL is actually deceptively simple. Each neural network was associated with a single person of interest and trained to recognize only that person. In order to determine who produced an unknown dataset presented to URIEL the unknown dataset is challenged against all available neural networks. From this point one of several evaluation strategies is used to determine the final decision. Even small variations in the decision making process that do not increase the number of incorrect decisions rendered can have a significant effect on how often URIEL chooses to render a decision at all thereby affecting overall system performance.

3.1.1. King of the Hill

When URIEL utilizes the King Of the Hill (KOH) decision making strategy the output of the internal neural nets are considered in a very simple manner. The strongest positive (greater than zero) response from one of URIEL's internal neural nets is considered the winner. In the event that none of URIEL's internal neural nets responds in a positive manner no neural net is considered the winner and URIEL will refuse to render a decision. This decision making strategy represents a more even balance between risk and the willingness of URIEL to render a decision.

3.1.2. King of the Hill with Threshold

URIEL may utilize a variant of the KOH decision making strategy wherein an arbitrary threshold is set and the KOH winner must have a response stronger than the arbitrary threshold value in order to be declared the winner. In this manner a specific confidence level can be guaranteed for all URIEL rendered decisions. In practice this serves to make URIEL behave in a more skeptical manner and reduces the instances where URIEL will be willing to render a decision favoring "confidence in" over "quantity of" decisions.

3.1.3. Least Hated

The Least Hated (LH) decision strategy is a unique decision strategy available for use with URIEL in the sense that it is the only one that will allow a decision to be rendered even when all of URIEL's internal neural networks

have extremely low levels of confidence. It functions as a simple decider: on a scale of negative to positive one the neural net that responds with the value closest to one, even if negative, wins. When URIEL employs this decision making strategy a decision will always be rendered regardless of overall confidence.

3.1.4. Least Hated with Threshold

Very similar to the Least Hated URIEL decision making strategy, Least Hated with Threshold (LHT) functions as a slightly less promiscuous LH strategy. This decision making strategy is particularly useful for scenarios when a decision must be rendered in most circumstances, LH does not provide a sufficient level of confidence in the decisions being made, and KOH is unable to render a decision in a timely fashion.

3.2. Comparison of URIEL Decision Making Strategies

In order to identify the most appropriate decision making strategy for a variety of situations the effects of interchangeable decisions making strategies upon both pure accuracy and fail secure acceptable performance were measured.

Table 1. Performance of Decision Strategies

Decision Strategy	Percent Correct	FSA (percent)
King of the Hill	82.3	98.0
King of the Hill with Threshold	16.0	98.0
Least Hated	93.1	93.1
Least Hated with Threshold	90.1	97.0

As can be seen in the table above depicting decision making performance, the decision strategies available have very different levels of usability for this system. The first column of the table measures how often a correct decision is rendered vs. all other possible outcomes including an incorrect decision and a refusal to render a decision. Contrasting with what is shown in the first column the second column, labeled FSA, shows the level of successful operation assuming that failing secure is acceptable. To clarify, the FSA column above shows instances wherein URIEL either rendered a correct decision or refused to render a decision based upon decision making strategy in use.

The KOH decision making strategy performs above expectations rendering a correct decision approximately eighty-two percent of the time while limiting incorrect decisions to less than two percent of all outcomes. This is considered to be an excellent level of performance since URIEL will render correct decisions more than eighty percent of the time while still not rendering incorrect decisions more than two percent of the time. From a usability perspective less than a eighteen percent occurrence of a refusal to render a decision is considered acceptable as this only mildly hinders the time responsiveness of the system when a decision is requested.

KOHT is notable in that the number of instances when a decision was rendered was drastically reduced. This is in line with the expectation that a stricter confidence requirement would result in fewer decisions rendered. KOHT maintains the same level of incorrect decisions rendered at just under a two percent occurrence rate.

The LH strategy enables URIEL to take additional risks resulting in a guaranteed decision. It is worth noting that during testing the LH strategy actually enabled the largest number of correct decisions, approximately a ten percent increase over KOH for an accuracy rating of ninety-three percent. As expected, this strategy also increased the number of incorrect decisions rendered relative to KOH by just fewer than five percent for a percentage chance to render an incorrect decision of just fewer than seven percent. This strategy is very useful for situations where the highest levels of accuracy are not required, a seven percent error tolerance is acceptable, and rapid decision making is required since it forces decisions to be made.

Finally, the LHT decision making strategy allows URIEL to make decisions with levels of confidence too low for KOH while still enforcing a minimal level of confidence in the decisions rendered. This decision making strategy exceeds both KOH and KOHT in its ability to render an accurate decision but renders a decision accurately three percent less often than the LH decision making strategy. Where this strategy truly excels is when its ability to limit

insecure failures (actually choosing wrongly), while slightly less effective in this regard than either KOH or KOHT this decision making strategy is within one percentage point of their reliability exceeding LH by approximately four percent for an overall reliability, in regards to the FSA metric, of approximately ninety-seven percent.

4. Results and Conclusions

This study has proven that behavioral biometric techniques can be used to extract meaningful information from mundane computer activities unobtrusively and to significant effect. Additionally this data has proven easy to process and shows strong potential for continuous authentication usages.

The abilities of neural networks to process and extract patterns useful for the identification and classification of humans from behavioral biometric data have been confirmed by this study. The level of accuracy URIEL was able to achieve is a strong indicator of the potential artificial intelligence techniques have in the interpretation of behavioral biometric data.

Finally this study has confirmed that behavioral biometrics data gathered over time can be used to allow artificial intelligence techniques to continue to learn even as input data from humans fluctuates over time. This continuous learning is the key in which long term profiles for humans can be developed and maintained allowing a constant level of accuracy in both identifications and classifications even over long durations of time.

This methodology, the KIT based behavioral biometric, implementation of URIEL, and even the architecture of the DEM can be used either directly or indirectly in creating real-world constructs. The constructs could, in turn, be used to ensure security, discover the identity of users, and mitigate threats from social engineering attacks against production systems on sensitive networks or resources.

In conclusion through URIEL and the methodologies presented in this research accurate, reliable, and quick decisions can be rendered using nothing more than a user's natural typing patterns. Forgoing advanced equipment, expensive hardware, or computationally expensive operations the URIEL platform and derivatives represent an enormous opportunity for the improvement of computer security. The methodologies and implementations presented in this research should prove to be solid components with which to build a more secure operating system and potentially protect money, secrets, or even lives.

References

1. Serwadda, Abdul, Vir V Phoha, and Ankunda Kiremire. "Using Global Knowledge of Users' Typing Traits to Attack Keystroke Biometrics Templates." *thirtteenth ACM multimedia workshop on Multimedia and Security.* New York: ACM, 2011. 51-60.
2. Bergadano, F., D. Gunetti, and C. Picardi. "Identity verification through dynamic keystroke analysis." *Intelligent Data Analysis* (IOS Press Amsterdam) 7, no. 5 (October 2003): 469-496.
3. Gianvecchio, Steven , Zhenyu Wu, Mengjun Xie, and Haining Wang. "Battle of Botcraft: fighting bots in online games with human observational proofs." *Proceedings of the 16th ACM conference on Computer and communications security.* 2009. 256-268.
4. Ngugi, Benjamin, Beverly K Kahn, and Marilyn Tremaine. "Typing Biometrics: Impact of Human Learning on Performance Quality." *Journal of Data and Information Quality (JDIQ)* (ACM New York) 2, no. 2 (February 2011): 21.
5. Lee, Jae-Wook, Sung-Soon Choi, and Byung-Ro Moon. "An Evolutionary Keystroke Authentication Based upon Ellipsoidal Hypothesis Space." 9th annual conference on Genetic and evolutionary computation. New York: ACM, 2007. 2090-2097.